

# Oscillatory synchronization model of attention to moving objects

Ozgur Yilmaz \*

National Research Center for Magnetic Resonance (UMRAM), Bilkent Cyberpark Ankara, Turkey  
Department of Psychology, Bilkent University, Ankara, Turkey

## ARTICLE INFO

### Article history:

Received 9 August 2010

Received in revised form 26 December 2011

Accepted 27 January 2012

### Keywords:

Attention

Neural synchrony

Cortical oscillations

Object tracking

## ABSTRACT

The world is a dynamic environment hence it is important for the visual system to be able to deploy attention on moving objects and attentively track them. Psychophysical experiments indicate that processes of both attentional enhancement and inhibition are spatially focused on the moving objects; however the mechanisms of these processes are unknown. The studies indicate that the attentional selection of target objects is sustained via a feedforward-feedback loop in the visual cortical hierarchy and only the target objects are represented in attention-related areas. We suggest that feedback from the attention-related areas to early visual areas modulates the activity of neurons; establishes synchronization with respect to a common oscillatory signal for target items via excitatory feedback, and also establishes de-synchronization for distractor items via inhibitory feedback. A two layer computational neural network model with integrate-and-fire neurons is proposed and simulated for simple attentive tracking tasks. Consistent with previous modeling studies, we show that via temporal tagging of neural activity, distractors can be attentively suppressed from propagating to higher levels. However, simulations also suggest attentional enhancement of activity for distractors in the first layer which represents neural substrate dedicated for low level feature processing. Inspired by this enhancement mechanism, we developed a feature based object tracking algorithm with surround processing. Surround processing improved tracking performance by 57% in PETS 2001 dataset, via eliminating target features that are likely to suffer from faulty correspondence assignments.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

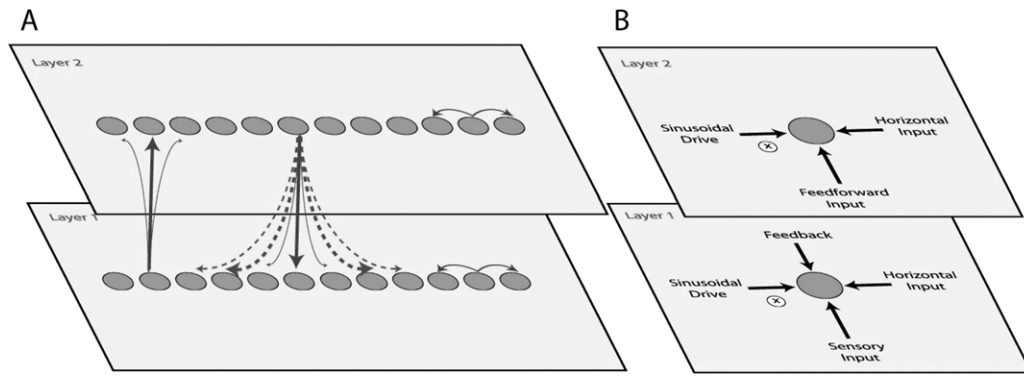
Visual attention organizes the deployment of limited computational resources, and selects the relevant stimulus from multiple stimuli for further processing. The neural representation of the attended stimulus in the brain becomes more salient compared to non-attended stimuli. One possibility is that the neurons subserving the processing of a stimulus respond more vigorously when the stimulus is attended compared to when it is not. However, substantial evidence show that influence of attention is not always achieved by an increase in the average firing rate of the neurons responding to the stimulus (Fries, Reynolds, Rorie, & Desimone, 2001; Luck, Chelazzi, Hillyard, & Desimone, 1997). Alternatively, it has been suggested that attention facilitates neural processing in the region of interest by increasing the temporal synchronization among the active neurons representing that region (for a review, see Engel, Fries, & Singer, 2001; Fries, 2009). In this view, neural synchrony increases the saliency of the selected region because correlated firing of a population of neurons has stronger impact

on a connected target population, compared to temporally incoherent firing (Abeles, 1982; Konig, Engel, & Singer, 1996). Therefore, the influence of attention extends to individual spikes and by modulating the timing of spikes, relevant stimuli are selected over irrelevant stimuli (Engel et al., 2001). Synchronization among a population of neurons can be achieved via phase synchronization of oscillatory firing patterns and this mechanism is shown to be one of the most energy efficient mechanisms (Buzsaki & Draguhn, 2004; Mirollo & Strogatz, 1990; Winfree, 1980). Experimental studies indicate the existence of such oscillatory patterns of activities along with synchronization among neural assemblies (Castelo-Branco, Neuenschwander, & Singer, 1998; Konig, Engel, Roelfsema, & Singer, 1995) in mammalian cortical areas. Recent evidence (for a review, Tiesinga, Fellous, & Sejnowski, 2008) indicates that phase of the neural firings with respect to an oscillatory signal determines which signals are propagated to higher areas in the neural hierarchy. However, it is still unclear how phase of oscillatory signals is related to visual attention. In this paper, we present a novel neural model and demonstrate that phase modulation can be achieved by inhibitory surround of feedback signals and it serves to reject distractor stimuli from attentional awareness as well as to enhance feature computation of distractors.

To study how oscillatory neural activity may relate to attention on moving objects, we propose a two layer computational neural

\* Correspondence to: Bilkent University Main Campus, Cyber Plaza, C Block, 2nd Floor, Bilkent, Ankara 06800, Turkey. Tel.: +90 312 290 1154.

E-mail address: [yilmazozgur81@yahoo.com](mailto:yilmazozgur81@yahoo.com).



**Fig. 1.** (A) The structural organization of the model. Various connections are illustrated sparsely and separately for clarity. Layer 1 sends excitatory signals to Layer 2 and activity from Layer 2 is sent to Layer 1 in an excitatory center-inhibitory surround manner. Also there are dynamic horizontal connections within Layer 1 and Layer 2 such that neighboring neurons excite each other based on the direction of stimulus motion. Solid (dashed) arrows indicate excitation (inhibition). The thickness of the arrow symbolizes the strength of the connection. There is only one spatial dimension in the network for computational economy and simplicity of presentation. (B) Inputs to a neuron in Layer 1 and Layer 2. Layer 1 neurons receive sensory input, feedback from neurons in Layer 2, modulatory input from neighboring neurons in Layer 1 and a sinusoidal drive signal. Intralayer connections and connections from Layer 2 modulate the gain of sensory input. The sinusoidal drive signal (common clock) multiplies the sensory input causing oscillations in neuronal firing. Layer 2 neurons receive input from Layer 1, input from neighboring neurons in Layer 2 and a sinusoidal drive signal. Similar to Layer 1, the same sinusoidal drive multiplies the primary input (Layer 1's output) producing oscillatory firing.

network model composed of integrate and fire neurons with synchronization capabilities (Fig. 1(A)). The neural network study in this paper is not an attempt to provide a complete model for Multiple Object Tracking (MOT) paradigm, but object tracking is used as an example to illustrate attentional phase modulation and surround suppression. Layer 1 in the model corresponds to a lumped representation of the low-level visual areas assigned for pre-processing that encode the location and low-level features of multiple objects, while Layer 2 corresponds to attention-related areas in which only attended objects are primarily represented. The anatomical locus for the Layer 2 is unclear and may include the terminal cortical areas of the dorsal and the ventral streams. We assume that both layers are retinotopically organized (for evidence of attentional retinotopic areas see Colby & Goldberg, 1999; Silver, Ress, & Heeger, 2005; Thompson & Bichot, 2004). Hence, activity in a given Layer 2 neuron suggests attentional deployment to the corresponding retinal location. The connectivity from Layer 1 to Layer 2 is on-center type while that from Layer 2 to Layer 1 is on-center-off-surround type. There are lateral excitatory connections within each layer that are dynamically changed to produce a predictive modulation of activity in the direction of motion.

There are three original contributions of our model on attentional modeling:

1. The modeling studies in the literature focused on attended stimuli, the main focus of our model is representation and processing of distractor stimuli.
2. It is an anatomically inspired hierarchical model with two layers and feedforward/feedback connections. The computational models of neural synchrony for attentional processing in the literature (see Tiesinga et al., 2008) have detailed mechanisms for individual neurons but most of the time simple network structures such as single layer feedforward network is simulated (an exception is Ardid, Wang, Gomez-Cabrero, & Compte, 2010).
3. Low level visual feature computation of distractor stimuli in the vicinity of attended region is enhanced in our model, which is suggested to improve tracking performance. The attentional mechanisms proposed in this paper can be applied to computer vision, specifically object tracking algorithms. Feature based object tracking algorithm with surround processing is provided in Section 3.5.

The most recent evidences for the essential mechanisms in our model come from physiology and neural imaging studies and are summarized below:

1. Radman, Su, An, Parra, and Bikson (2007) investigated the effect of Local Field Potentials (LFPs) on neural spikes and found that hyperpolarizing LFP delayed action potential while depolarizing LFP advanced it. In our model, an oscillatory signal representing LFP modulates the phase of the neural spikes.
2. Fries (2009), Mitchell, Sundberg, and Reynolds (2009) and Womelsdorf et al. (2007) studies suggested that rhythmic activity in the network causes rhythmic modulation of excitatory synaptic input gain of the neurons. Also, studies of Lampl and colleagues Lampl and Yarom (1993); Lampl et al. (1999) suggest a multiplicative interaction between LFP and synaptic summation. Therefore, the oscillatory LFP signal in our model multiplicatively modulates the input gain of the neuron, which is novel to our knowledge. It should be noted that multiplication is suggested to be one of the most common non-linear operations realized in the nervous system (Schlotterer, 1977; for a review, Koch & Segev, 2000). For generating oscillatory firing, an external modulation signal is used in the network instead of a network interconnected inhibitory and excitatory neurons. This simplification is validated by the findings in literature mentioned above and it is necessary for studying attentional mechanisms in a complex network with feedforward and feedback connections.
3. It is shown that phase of the neural activity with respect to LFP determines which signals are transmitted up in the visual processing hierarchy (review Tiesinga et al., 2008). And LFP of distant neural populations can be correlated (Siapas, Lubenov, & Wilson, 2005). In our model, feature encoding layer and attentional/saliency map share the same LFP drive that allows for filtering out the distractor neural activity.
4. Sundberg, Mitchell, and Reynolds (2009) showed that attentional surround suppression is more delayed compared to enhancement. The surround feedback connections from attentional/saliency map to feature encoding layer is more delayed in our model.
5. Maier et al. (2008) showed that V1 activity may survive even for perceptually suppressed stimuli. Therefore, low level visual features may be computed even though the stimulus is not consciously registered. In our model, attentional surround triggers activity in the first layer of the network hence initiates low level feature processing however this activity does not propagate to second layer (and not attentively tracked). We are suggesting that distractor activity in low level feature processing neural substrate is essential for the successful tracking of target objects.

6. Kamondi et al.'s (1998) neural model successfully explains the phase coding in hippocampus. Kamondi et al.'s model and the model presented in this paper show similarities in behavior, although the amount of detail for single neuron in their model is much higher than ours (see Discussion for a detailed comparison). Also, Koepsell and Sommer's (2008) multiplicative model which is very similar to the model presented in this study successfully predicts the activity in cat LGN and shows that "oscillations significantly contribute to the information carried in the spike train".

We propose that attentional surround not only modulates the phase of the neural activity for preventing distractors reaching visual awareness but also enhances low level visual computation of distractors in order to avoid faulty correspondence assignments of target features. Inspired by our neural network model, we developed a novel object tracking algorithm (Section 3.5) in which surround processing substantially improves the target tracking performance. In our object tracking algorithm, the features in the attentional surround are processed in order to improve feature processing in target region.

In Section 2, neural network and the object tracking algorithm methods are given. In Section 3 simulations of the neural network model for single target, multiple targets and Pylyshyn's (2006) probe detection experiment are presented. The novel object tracking algorithm and tracking results on PETS 2001 dataset are presented in Section 3.5. Discussion section provides detailed evidence for the proposed mechanisms of the neural network model and suggestions for future work.

## 2. Methods

### 2.1. Neural network model

The neurons in both layers are leaky integrate-and-fire neurons with previously tested cell parameters (parameters imported from Brody & Hopfield, 2003). The integrate-and-fire model neurons integrate the input applied to them while leaking some portion of the activity at each time step. If the activity exceeds a certain threshold it generates a neural spike. Immediately after a spike, there is a brief refractory period during which the neuron cannot integrate its input.

The inputs to each neuron in both layers are shown in Fig. 1(B). Layer 1 neurons receive sensory inputs, inputs from Layer 2 and lateral excitatory inputs from neighboring neurons. The gain of the sensory input is modulated by the output of Layer 2 and Layer 1 lateral inputs. Therefore Layer 2's output and Layer 1 lateral connections modulate but do not initiate neural activity in Layer 1 (Hupé et al., 1998). A sinusoidal signal multiplies the sensory input in Layer 1 (see Fig. 1(B)) and is expected to produce oscillatory behavior in neuronal spiking (Ferster & Carandini, 1996; Koch & Segev, 2000; Koepsell & Sommer, 2008; Lampl & Yarom, 1993). We assume that an assembly of pacemaker cells serves as a basis for the sinusoidal signal used in the model. This sinusoidal signal (also see Discussion, Section 4.4) is used as a common synchronizing clock for both layers (Buzsáki & Chrobak, 1995; Fries, Roelfsema, Engel, König, & Singer, 1997; Hopfield, 1995). The period of the sinusoidal clock is 285 simulation time units, and it corresponds to gamma range (35 Hz) oscillations, assuming 10,000 time units is equivalent to one second. In Layer 2, neurons receive input from Layer 1 and lateral input from neighboring neurons and again the same sinusoidal drive multiplies the Layer 2 input and causes oscillations in the spiking pattern. The inhibitory signal from Layer 2 to Layer 1 is delayed compared to the excitatory signal from Layer 2 to Layer 1. The signals in the intra-layer lateral connections in Layer 2 are delayed compared to all other input signals (Bringuier, Chavane,

Glaeser, & Fregnac, 1999; Grinvald, Lieke, Frostig, & Hildesheim, 1994). These delays are critical parameters in the model because the sinusoidal clock signal puts constraints on the time interval of integration of the neural activity.

The mathematical description of the model and its parameters are available in the Appendix A. We tested this model by presenting multiple moving stimuli at its input. In response to the simulated input, the spiking activity of each model neuron in the network was examined.

### 2.2. The object tracking algorithm with surround processing

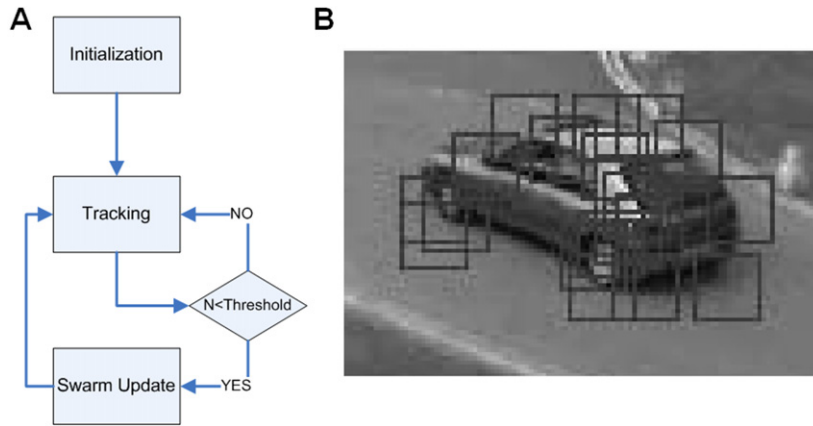
In order to illustrate the advantages of surround processing in attentional tracking of moving objects we present a swarm intelligence based feature tracking algorithm. Corners are widely used in computer vision for their stability and immunity to aperture problem (Shi & Tomasi, 1994). Cornerness is an analog entity depending on the difference between the two edge directions that constitute the corner, i.e. a strong corner when there is 90° angle between two high contrast edges. A patch of image in the neighborhood of a corner can be tracked robustly in a subsequent frame by finding the image patch that best correlates (pixel by pixel) with the original image patch in the reference frame. Therefore, an image patch is searched in the neighborhood based on pixel by pixel similarity (correlation). This technique is called normalized cross correlation, normalization being performed by subtracting the means of image patches from each patch and dividing with the standard deviations of the two image patches. The image patch (16 × 16 pixels in our algorithm) that is tracked in every frame is called a feature. Our tracking algorithm utilizes corner detection and normalized cross correlation, which are standard algorithms in object tracking literature. The tracking algorithm and surround processing are explained in two sections below and the details are given in Appendix B.

#### 2.2.1. Swarm of trackers

"Swarm intelligence is a property of systems of unintelligent agents of limited individual capabilities exhibiting collectively intelligent behavior" (White & Pagurek, 1998). Artificial swarm intelligence includes designing algorithms or distributed problem-solving devices inspired by the collective behavior of social insects and other animal societies (Bonabeau, Dorigo, & Théraulaz, 2000). In bird flocks, each bird obeys a small set of very simple rules based on the behavior of the other birds in the neighborhood, however very complex flock behavior emerges. Similarly in our algorithm, there is a swarm of tracking windows each of which performs very basic image processing computations on the images of a moving object, but the cumulative behavior of the swarm shows complexity. This approach allows robust tracking of sub-parts of a moving object, hence is immune to size and appearance changes in the moving object's images. The algorithm initiates small correlation tracking windows on the corners of the target (phase congruency corners, Kovess, 1999, 2003) and the cumulative motion of the tracking windows determines the behavior of the swarm. The swarm is refreshed aperiodically to cope with the changes in the target shape. The position and the size changes of the target are decided based on both the motion of each small tracking window and the overall spread of swarm members. Thus, both dynamic and geometric cues are used to determine changes in size of the target. Under extreme target motion on the image, overall target window size is automatically increased in order to keep the track, which is adjusted back as the motion stabilizes.

The algorithm has 3 phases (Fig. 2(A)): initialization, tracking and swarm update. In initialization phase, a certain number normalized cross correlation tracking windows (image patches)





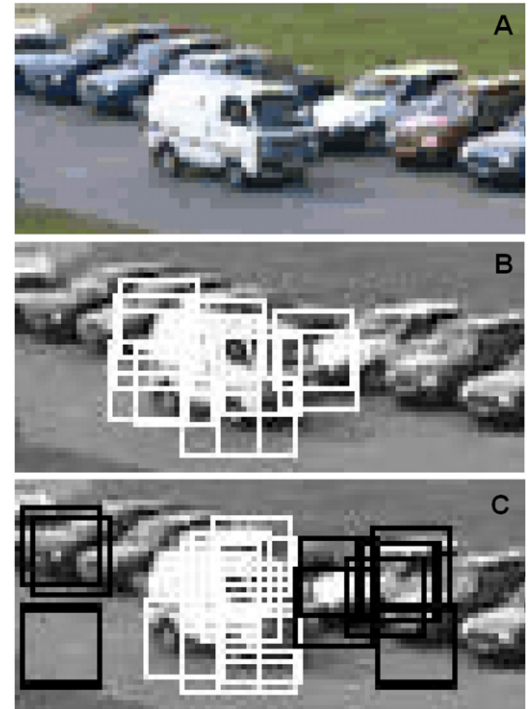
**Fig. 2.** (A) The flowchart for the three phases of the object tracking algorithm. After Initialization of the tracking windows, Tracking phase starts. When the number of tracking windows falls below a threshold, swarm update phase is initiated in which the tracking windows are reselected. (B)  $16 \times 16$  pixels tracking windows on the target corners are shown. These tracking windows are initiated in initialization and swarm update phases and tracked by normalized cross correlation in tracking phase.

are initiated on strong corners of initial image of the moving object (Fig. 2(B)). In tracking phase, these windows are tracked using normalized cross correlation and the motion of these windows is used (statistically) in the calculation of moving object's location and size. The distribution of the pixel motion of tracking windows is calculated at each image frame. This statistic generates an estimate of collective motion in the image frames (swarm motion), which is due to the actual motion of the moving object. Using this statistic, the algorithm is able to reject outlier tracking window motion (faulty correspondences) and arrives at a robust estimate of object motion. The tracking windows which have motion inconsistent with the swarm motion are considered as outliers and dropped, hence swarm is refined with every analyzed image frame. The object's motion is calculated from the mean motion of inlier tracking windows. If the total number of tracking windows falls below a threshold, swarm is re-spread on the target area defined by the center of the swarm and the size of the target in the swarm update phase. The details of the algorithm are provided in Appendix B.

### 2.2.2. Surround processing

The simulations of the neural network presented here suggest that the feature processing for distractors in the surround of tracked targets is enhanced (Results, Section 3.4). This enhancement can be beneficial for tracking performance such that the features of targets that show resemblance to features of distractors can be suppressed to avoid faulty correspondence assignments. Therefore we suggest that, one of the purposes of surround processing is to reduce the probability of faulty correspondence assignments during tracking of moving object's features. This is achieved by the following steps in tracking algorithm:

1. Initialize tracking windows on the target object's strong corners as explained above.
2. Compute the corners in target object's surround and extract correlation windows (features) at the strong corner locations of the surround (Fig. 3(C), black squares).
3. Calculate similarity metrics between target object's features and the features in the surround.
4. Assess the quality of target object's features based on their similarity and Euclidean distance with surround features.
5. Select a subset of target object's features with highest quality in terms of dissimilarity with surround features (Fig. 3(C), white squares).



**Fig. 3.** (A) Target region in the test image of PETS 2001 dataset. (B) Features (tracking windows white squares, only 12 of them for illustration purposes) are initialized in the corners of the target in the case of no surround processing. (C) Surround features (black squares) and target features (white squares) are shown. Surround features are used to select the best target features that have minimum correspondence ambiguity problem.

In our algorithm, surround is defined as the region around a corner that extends beyond the  $16 \times 16$  pixels of target region. Eliminating the initial target features (strong corner features) that are similar to surround features is expected to improve the normalized cross correlation tracking by sharpening the correlation distribution over space for each tracked feature. Having a sharper correlation distribution leads to less error in tracking and less number of false correspondence assignments. The improvement of tracking with surround processing is demonstrated in Results section. The distractor features are not included in the computation of target object's motion; they are only used for quality assessment of the target features. The details of the surround processing are given in Appendix B.

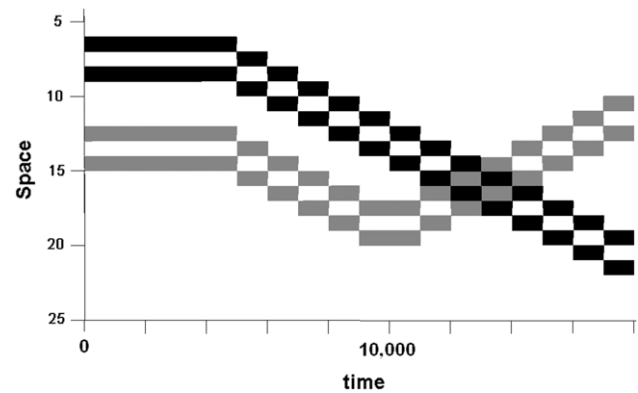
### 3. Results

Neural network simulations and object tracking results on PETS 2001 dataset is presented below. The simulations illustrate the phase modulation of neural activity (Fig. 5) and enhancement for computation of distractor features in the attentional surround (Fig. 11).

#### 3.1. Suppression of distractors via temporal tagging

A simple tracking simulation is demonstrated with the model by presenting two moving objects (circular annuli) as sensory inputs. The network contains only one spatial dimension for computational simplicity and ease of understanding. Thus the stimuli used in simulations are 1D projections of the circular annuli (see Fig. 4). The time scale of the model's output is not calibrated and does not quantitatively follow the physiological data in the literature. The target that has to be tracked is identified by changing the saliency (e.g. brightness) of the target for a brief period of time before the objects start to move. To simulate this tagging process, the target object's saliency was increased for a period of time (1000 simulation time units) in the beginning by increasing the sensory input to cells in Layer 1 that represent the annulus designated as the target. After tagging, one of the objects is designated as target (upper black object in Fig. 4) and the other is designated as distractor (lower gray object in Fig. 4). The tagging process produces a transient increase in activity in Layer 2 and hence creates an "attentional window" for the target object. This transient increase is due to the gain characteristics of feedforward connections such that a gain larger than a threshold produces activity in Layer 2. Then Layer 2 activity reinforces itself with positive feedback, which maintains an "attentional window" at the location of tagging. Another mechanism to initiate activity in Layer 2 is to increase Layer 2 neurons' gains, possibly due to volition. While tagging of Layer 1 neurons can be thought as exogenous, increasing Layer 2 neuron gain is endogenous. Therefore, in order to initiate activity in Layer 2, either feedforward input should be elevated or Layer 2 neurons should be excited, otherwise the default mode of operation is to filter out every feedforward input. At  $t = 5000$ , both objects start to move, and one of them reverses motion direction at  $t = 10,000$ . They overlap around  $t = 12,000$ . The simulation ends at  $t = 18,000$ .

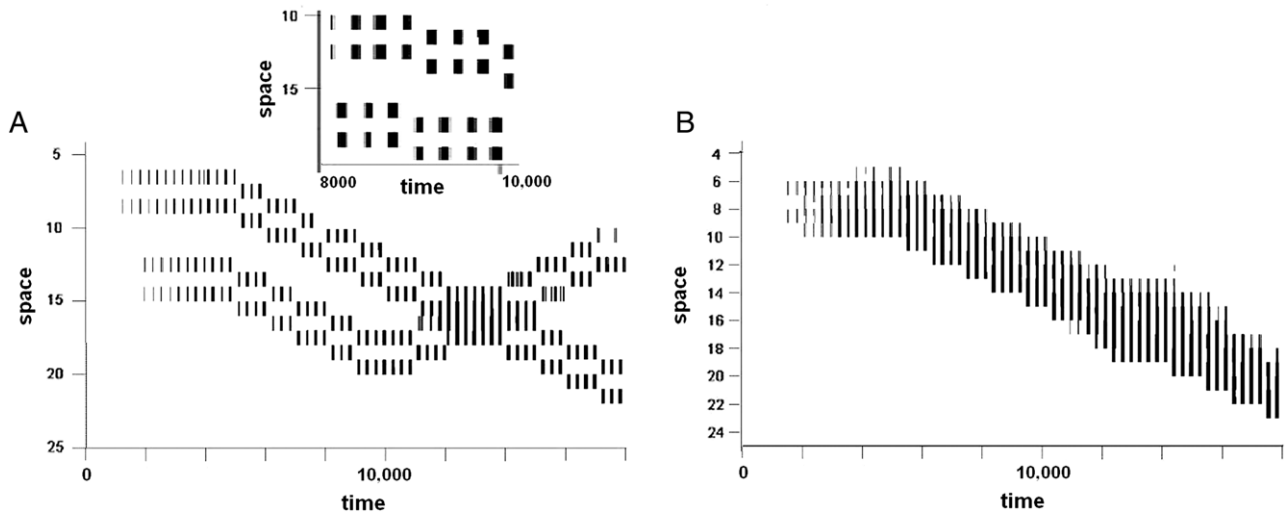
Spiking activities of neurons in Layer 1 and Layer 2 are shown in Fig. 5(A) and (B) respectively. The black bars in the space–time diagram represent spikes, single spikes are not resolvable but the width of the bars is a measure of the number of spikes during the burst. In Layer 1, the neurons fire in oscillatory bursts of spikes and the frequency of bursts is the same as the clock frequency. The neurons that receive input from the target have bursts that are in synchrony with the clock: the spikes overlap with the positive half of the cycle. However, the neurons that receive input from the distractor have bursts that are  $180^\circ$  out-of-phase with the clock: the bursts overlap with the negative half of the sinusoidal cycle (see the inset in Fig. 5(A) for the out of phase firing of the target and the distractor). The reason for this out-of-phase firing is the inhibitory signals received from Layer 2. When a neuron in Layer 1 receives a large inhibitory signal, its net gain becomes negative, causing the neuron to integrate during the negative cycles of the clock. The possible neural mechanisms for integration of input during negative cycles are explained in the Discussion section. The integration during the negative cycles of the clock causes out-of-phase activity of the neuron with respect to the phase of the clock signal. The end result is that the neurons representing the distractor are tagged in Layer 1 via feedback inhibition. Because the firing of distractor neurons in Layer 1 is  $180^\circ$  out-of-phase with the clock, the signals from these neurons sent to Layer 2 do not



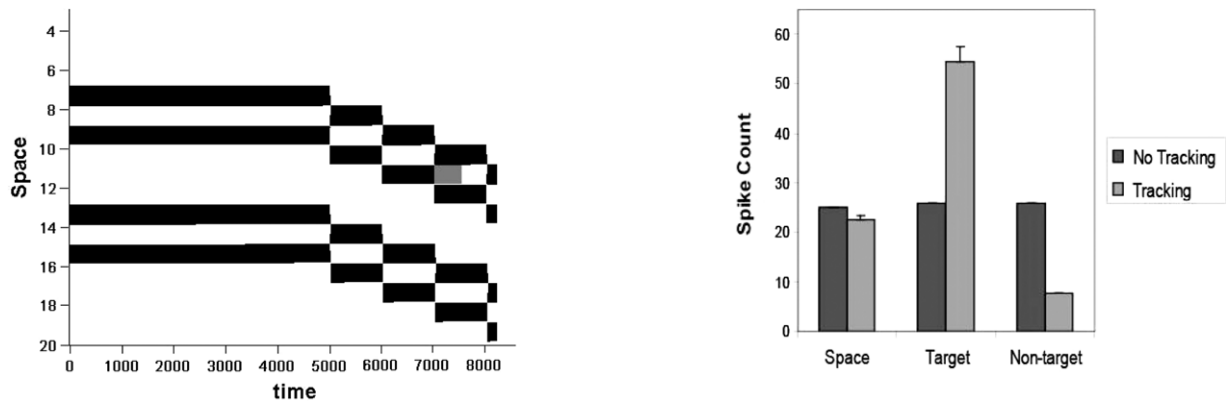
**Fig. 4.** The sensory input applied to Layer 1 for the first two simulations of the model. The x axis is time and the y axis is space. The y-axis also represents the neuron number in the retinotopically organized set of neurons. One dimensional projections of two circular annuli are presented as input to the model (magnitude of the input is 1). Upper black object is the target and the lower gray object is the distractor. The amplitude of the sensory input is identical for the two objects, thus the gray level of the objects in the figure does not represent saliency, but they are used for presentation purposes only. The target is identified by increasing its saliency (magnitude set to 4) in the [1000, 2000] time interval. The objects start to move at  $t = 5000$  and the distractor reverses motion direction at  $t = 10,000$ . They overlap at around  $t = 12,000$ .

coincide with the positive cycles of the clock. In Layer 2, the clock signal is multiplied with the signals coming from Layer 1, hence the Layer 1 distractor neurons become ineffective in activating Layer 2 neurons (Fig. 5(B)). However, the target neurons in Layer 1 fire in synchrony with the clock hence are effective in exciting Layer 2 neurons. In summary, both target and distractor produce activity in Layer 1, but only the target largely activates neurons in Layer 2 and thus the activity in Layer 2 represents attentive tracking of the target (Jovicich et al., 2001) and filtering of the distractor stimuli proximal to the target. It should be noted that distractor stimuli far away from the target stimuli may require other filtering mechanisms. Also note that this simulation uses non-overlapping objects in one spatial dimension; however in other simulations (not reported here) we have used objects that overlap and occlude each other, and these simulations yield results similar to those shown in Fig. 5.

A probe detection experiment (Fig. 19 in Pylyshyn, 2006) is simulated by presenting a probe flash at the center of the target or distractor or in the background space region. In separate tests of the model, the flash is presented at the locations of target (location = 11, Fig. 6), distractor (location = 17) or empty region (location = 3) at  $t = 7000$ , for 500 simulation time units. The uncalibrated luminance value of the flash was one fifth of that of the moving objects. The visibility of the flash is estimated by counting the spikes in Layer 1 neuron at the spatial location where the flash is presented. The integration time over which the spikes are counted is 1000 simulation time units (from  $t = 7000$  to  $t = 8000$ ). When the probe is presented on the background space region, its location is chosen (location = 3) to be the same distance from the target object as the distractor probe flash location (location = 15), so that the feedback inhibition on the flashes due to the target object are identical. There are two conditions as in the case of Pylyshyn's experiment: Tracking and No Tracking. The original two layered model is simulated for the Tracking condition, but the feedback connections from the second layer to the first layer are removed for the No Tracking condition since attentional tracking mechanisms are assumed not to operate in free viewing. The bypass of the network in the second layer during free viewing is in accordance with the task dependent processing in the brain (Bedell, Chung, Ogmen, & Patel, 2003), and it can be achieved by task dependent cognitive factors which are not modeled here. In



**Fig. 5.** (A) Spiking pattern of neurons in Layer 1 during attentive tracking of a single object (from  $t = 0$  to  $t = 18,000$ ). In this and other figures, a dark spot in the figure represents an action potential of the neuron residing at that location. A thick bar represents a burst of firing in a set of neurons in the same time window. The neurons show oscillatory bursts of spikes and the frequency of the bursts match the clock frequency. The neurons responding to the target (target neurons) produce bursts of spikes that are synchronous with the clock, whereas the neurons responding to the distractor (distractor neurons) produce bursts of spikes that are 180° out-of-phase with the clock. The inset shows out-of-phase firing of neurons responding to target and distractor more clearly. Note that both the target and the distractor trajectories are represented in Layer 1. (B) Spiking pattern of neurons in Layer 2 during attentive tracking of a single object. The neurons receiving inputs from target neurons in Layer 1 produce bursts of spikes that are synchronous with the clock, while the neurons receiving inputs from distractor neurons in Layer 1 do not fire at all. Thus, in Layer 2, only the target's trajectory is represented, whereas the representation of the distractor object is suppressed.



**Fig. 6.** The sensory input for the second simulation of the model. A flash was presented at the center of the circular target. This probe flash was presented at  $t = 7000$  for 500 time units. The luminance of the flash was one fifth of the luminance of the moving objects. The simulation was repeated for flashes superimposed at the distractor location and also in the empty background region.

order to compensate for the lack of positive feedback on Layer 1, sensory gain is increased in Layer 1 for No Tracking simulations. The simulations are repeated for several times and the spike counts at the flash locations are averaged.

The results of the simulations show that there is a large decrement in spike count when the flash occurred at the center of the distractor compared to at the center of the target during the tracking task (Fig. 7, Target vs. Non-target bars).

A sigmoidal function was implemented to transform the spike count into percent correct of the probe flash detection task. The equation of the function was:

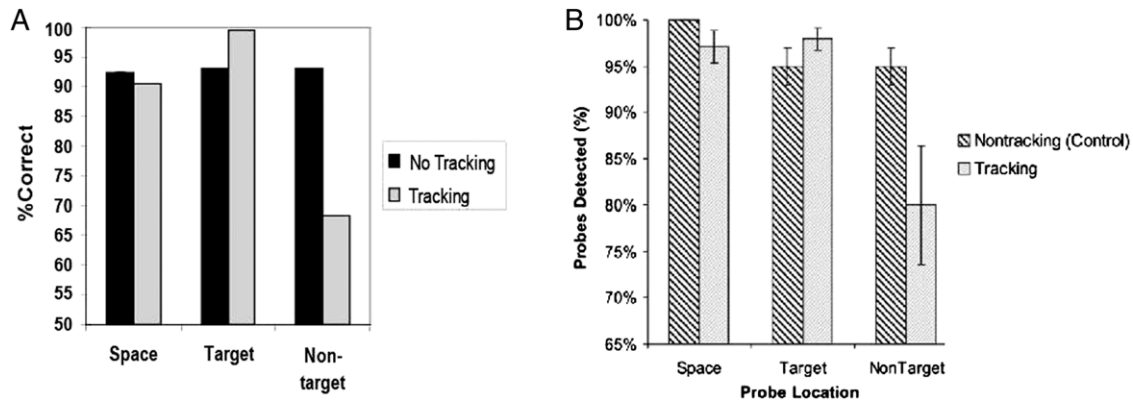
$$\text{Percent Correct} = 1/[1 + \exp(-0.1 * \text{SpikeCount})].$$

The parameter of the sigmoidal function is chosen to make the simulation and experimental percent correct results most similar especially for Non-target case, however the qualitative similarity of simulation and experimental results is independent of this parameter. The computed percent correct responses are shown in Fig. 8(A). During the tracking task, the visibility of a flash at a non-target (distractor) location is reduced compared to visibility

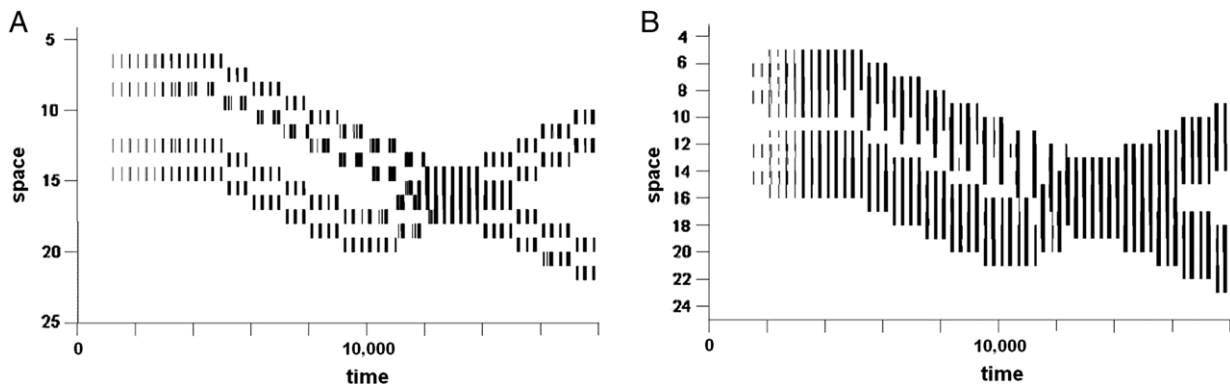
**Fig. 7.** The results of the simulation of the model. Average spike counts in Layer 1 at the target, distractor (non-target) and background (space) regions for Tracking and No Tracking conditions.

at the target location which is in excellent qualitative agreement with the experimental findings of Pylyshyn's (2006) (Fig. 8(B)). The mechanism in the model responsible for this visibility reduction is as follows. Before the presentation of the flash, distractor neurons produce bursts of spikes 180° out of phase with the clock. When the flash is presented in the gap at the distractor location (gray region in Fig. 6), the distractor neurons interfere with the activity of the neuron representing the flash's location via horizontal connections. This interfering activity from the distractor neurons is also 180° out of phase with the clock hence the effective gain of the neuron at the flash's location is reduced (see Appendix A). This reduction in the effective sensory gain reduces the number of spikes produced in the neuron at the flash's location in Layer 1.

Comparing Fig. 8(A) and (B) indicates a few quantitative differences: the detection rate of the probe flashed in background space is largest in the psychophysical experiment but it is not the case in the simulation. This difference might originate from the lack of masking interactions in the model from moving objects to the probe flashes. When the probe flash is presented close to the moving objects (both Target and Distractor), a mechanism of masking is expected to reduce the visibility of the probe (Ogmen, Breitmeyer, & Melvin, 2003); but this mechanism is not included



**Fig. 8.** (A) The computed percent correct responses of the detection of the probed flash task. The spike count (Fig. 4) was transformed into percent correct with equation provided in the text above. (B). The experimental data adapted from Pylyshyn's (2006).



**Fig. 9.** (A) Spiking pattern of neurons in Layer 1 during attentive tracking of two objects (from  $t = 0$  to  $t = 18,000$ ). The sensory input shown in Fig. 4 is used for this simulation. However, both the moving objects are assigned as targets in this simulation. In this case, the neurons responding to both the moving objects produce bursts of spikes that are synchronous with the clock. (B) Spiking pattern of neurons in Layer 2 during attentive tracking of two objects. The neurons receiving inputs from target neurons in Layer 1 produce burst of spikes that are synchronous with the clock. Both the moving objects are represented in Layer 2, indicating that they are both attentively tracked. Note the weakening of the neural activity for the upper object in the [5000, 12,000] time period when the two objects are close to crossing each other.

in the model thus the detection rate of the probe inside Target is larger than the detection of the probe in Space. Also note that, the large peak of spike count when probe flash was presented at the center of the target during tracking is trimmed down with the nonlinearity of the sigmoidal function.

### 3.2. Tracking multiple targets

In order to test the ability of the model to track multiple objects, the sensory input shown in Fig. 4 is used in another simulation. Instead of tagging only one object as in the previous simulation in this simulation, both of the objects were tagged in the beginning and assigned as targets. Spiking of neurons in Layer 1 and Layer 2 is shown in Fig. 9. Note that both of the objects are represented in Layer 2, therefore the model can adequately explain simultaneous attentive tracking of multiple objects. It should be noted that horizontal connection gain was biased in the direction of target motion in order to distinguish between a target and a distractor after their overlapping. Therefore it was assumed that motion direction was estimated and it was used to enhance activity propagation in the direction of target motion. This is consistent with motion anticipation theory (Yilmaz, Tripathy, Patel, & Ogmen, 2007).

### 3.3. A probable source for the capacity limit in attentive tracking of multiple objects

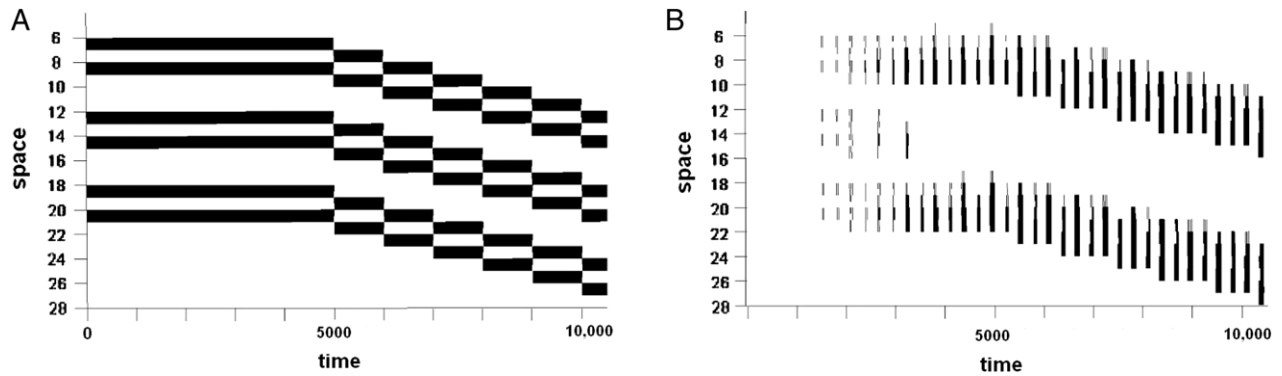
Weakened activity in Layer 2 for the upper object in [5000, 12,000] time period in Fig. 9(B) suggests interference between the activity produced by nearby target objects due to the

inhibitory surround from Layer 2 to Layer 1. This interference from the inhibitory surround might limit the capacity of attentive tracking in the model, as long exposure of input from the inhibitory surround from other targets may decrease the attentional focus on a target object. In a third simulation, three moving objects are used and all three of them were selected to be targets (Fig. 10(A)). As shown in Fig. 10(B), the inhibitory surround created by the uppermost and the lowermost moving objects prevented the emergence of activity corresponding to the middle moving object in Layer 2. The results show that local interactions among the target neurons may impair the tagging process and the ability of the model to attentively track multiple objects.

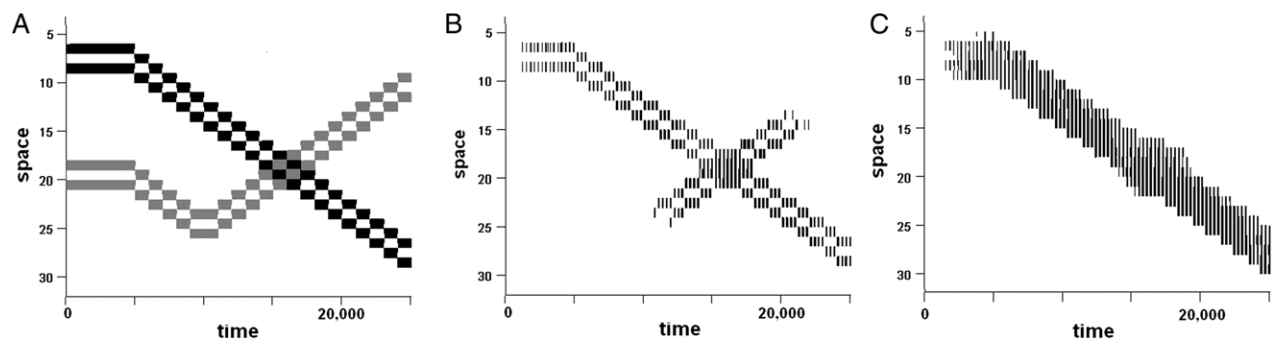
### 3.4. Effect of target–distractor separation

The fourth simulation investigated the effect of target–distractor distance on the neural activities in both layers. The target (black) and the distractor (gray) objects were more distant from each other compared to the previous simulations (Fig. 11(A)). Spiking of neurons in Layer 1 and Layer 2 are given in Fig. 11(B) and (C) respectively. In Layer 1, the neurons that receive input from the target object produce activity synchronous with the clock, while the neurons that receive input from the distractor object do not produce activity when it is distant from the target object. However, when the distractor comes close to the target and enters the inhibitory zone of the “attentional window”, distractor neurons produce bursts of spikes that are 180° out-of-phase with the clock (Fig. 11(B)). Thus the distractor object is largely represented in Layer 1 only when it is close enough to the target object. Similar





**Fig. 10.** (A) The sensory input for the third simulation of the model. One dimensional projections of three circular annuli are presented as inputs to the model (the magnitude of input is 1). All three objects are designated as targets and are expected to be attentively tracked by the model. The targets are tagged for attentive tracking by increasing their saliency in the [1000, 2000] time interval. The objects start to move at  $t = 5000$ . The simulation ends at  $t = 10,000$ . (B) Spiking pattern of neurons in Layer 2 during attentive tracking of three objects. Initially all three objects are represented in Layer 2, however the neural activity for the middle object ceases after a while due to inhibitory surrounds created by the two flanking targets.



**Fig. 11.** (A) Sensory input for the fourth simulation of the model. The input is very similar with the input for the first simulation (Fig. 4), except the starting locations of the target (black) and the distractor (gray) are more separated. The simulation ends at  $t = 25,000$ . (B) Spiking pattern of neurons in Layer 1 during attentive tracking of a target in the presence of a distant distractor. The neurons responding to the target produce bursts of spikes that are synchronous with the clock, whereas the neurons that respond to the distractor do not fire until the distractor comes close to the target and enters the suppressive zone of attention. The neurons responding to the distractor produce burst of spikes that are  $180^\circ$  out-of-phase with the clock as in the first simulation (Fig. 5(A)). (C) Spiking pattern of neurons in Layer 2 during attentive tracking of a target in the existence of a distant distractor. The neurons receiving inputs from target neurons in Layer 1 produce burst of spikes that are synchronous with the clock, while the neurons receiving inputs from distractor neurons in Layer 1 do not fire at all as in the first simulation (Fig. 5(B)).

to the previous simulations, only the target is represented in Layer 2 (Fig. 11(C)). In the model, Layer 1 corresponds to the pre-processing stage where the basic object features are represented. Therefore modeling results indicate that the distractor objects' features are largely pre-processed only when these objects come close to the target objects, otherwise low level features of the distractor objects are minimally processed. This is achieved by the surround of attention which enhances low level feature processing of distractor objects while preventing them from interfering with target objects (see Discussion, Oscillations and Neural Firing subsection for a discussion of this mechanism).

The lack of activity in Layer 1 (far away from target, outside the attentional surround) might be exaggerated in the model, however physiological studies suggest that there is a narrow time window of excitability for a neuron to accumulate synaptic input (Lampl et al., 1999; Lampl & Yarom, 1993). Hence, the neuron can be silent due to asynchrony. Winner-take-all type synchrony based processing that is guided by attention has been suggested to take place in the cortex (Börgers, Epstein, & Kopell, 2005; Fries, Nikolic, & Singer, 2007; Olufsen, Whittington, Camperi, & Kopell, 2003). In a very relevant study, Recanzone and Wurtz (2000) recorded from MT and MST neurons when the monkey attended moving stimuli. They have found that in the case of a long duration attention, the unattended stimuli do not interfere with the attended stimulus. And, distractor moving stimuli are expected to be filtered out as they go up to extrastriate cortex by a winner-take-all competition, if there is enough time for attention to build up. Neural recordings

from V4 suggest that neural responses are reduced more than half for an unattended stimulus (Luck et al., 1997; Moran & Desimone, 1985). However, the effect of attention is much smaller in V1 which suggests that attentional processing effectively suppresses neural representations of irrelevant stimuli after some level in the functional hierarchy. Therefore, the unattended objects might still have representations at the lowest levels of the visual processing but feedforward input to higher levels are effectively blocked. Chelazzi, Duncan, Miller, and Desimone (1998) recorded from IT cortex and found strong filtering of unattended stimulus. Everling, Tinsley, Gaffan, and Duncan (2002) recorded from prefrontal cortex neurons during a spatial attention task and reported strong, early and global filtering of unattended locations. These two studies support the idea that distractor representation gets weaker as it goes up in the visual stream. It should be noted that Layer 1 in the model is suggested to correspond to a retinotopically organized neural substrate hierarchically high enough so that attention can effectively suppress irrelevant information. The exact locus of this substrate is not specified.

None of the studies mentioned above investigated the neural activity of extrastriate neurons encoding the distractor stimulus, as a function of the distance to the target stimulus (Fig. 11). In Chelazzi et al. (1998), the activity of a distractor neuron is enhanced with the presence of a nearby target (Fig. 21 A, page 2936, Target = Poor stimulus alone vs. Target = Poor Stim. in 2-Stim Array), but this effect is not systematically examined in the experiments. We predict that the distractor neural activity will get



enhanced with decreasing distance between target and distractor and further physiological studies are needed to test this prediction.

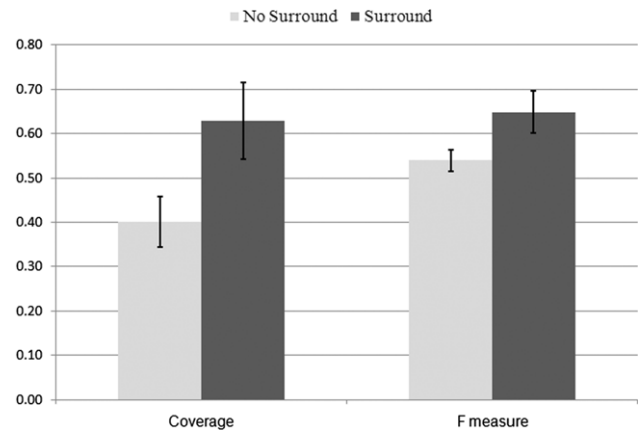
### 3.5. Surround processing in the object tracking algorithm

The model predicts enhancement of feature processing for distractor stimuli as they get closer to the target. This mechanism is energy saving: the features in the surround that have immediate possibility of causing correspondence mistakes are largely processed. In addition to efficiency, processing of surround features can reduce the correspondence errors by suppressing the target features similar to surround features. Since suppressed target features are more likely to make correspondence mistakes due to similar distractor feature in the immediate surround, the tracking performance is expected to increase. The surround processing algorithm proposed in Section 2.2.2 requires similarity calculation of target and surround features, and suppression of low quality target features. This can be achieved by cortical lateral inhibition mechanism in a spatiotopically organized network in which neurons encoding similar features reside in the same neighborhood (e.g. orientation maps).

It should be noted that the neural network simulations inspired the surround processing in object tracking but the connection with the neural network is not strong. Swarm of Trackers object tracking algorithm is executed with and without surround processing for PETS 2001 dataset (test set, camera 1 and 2). Seven object tracks with partial occlusions and relatively complicated surrounds are selected. For some videos there is a moving car in front of a parking lot. In some other videos there are people walking in front of moving or parking cars. Three of these tracks overlap with each other, hence there are distractor moving objects for some object tracks. Two sample videos are given in supplementary materials, that have moving distractor objects in the surround. The supplementary video also shows how the tracking fails for a moving van when there are many nonmoving cars in the background, and how surround processing improves tracking by rejecting features from the upper half of the target region since they are similar to background features. Precision and Recall statistics are derived based on intersection between hypothesized and correct object regions (see Yilmaz, Javed, & Shah, 2006 for a review of object tracking approaches and performance metrics). Coverage is a measure that is used to evaluate if the target object is being tracked. In coverage test,  $F$  measure ( $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$ ) should exceed a threshold to declare that target tracking is acceptable in a frame. For the tests,  $F$  measure threshold is selected as 0.5. The coverage (ratio of tracked frames) and the mean  $F$  measure with and without surround processing are given in Fig. 12. Surround processing improved coverage by 57% ( $t$  test,  $t(6) = 2.46$ ,  $p = 0.048$ ) and  $F$  measure by 20% ( $t$  test,  $t(6) = 2.51$ ,  $p = 0.039$ ). The improvement is attributable to suppressing the target corners that have the potential to be wrongly matched with surround corners, especially in cluttered tracking environments (see Fig. 3(C)).

## 4. Discussion

The main aim of the neural modeling is to understand how an oscillatory neural synchrony based model can provide a framework for studying attention to moving objects. In this study, a novel model is proposed which can attentively track a small number of moving objects. The model is a two layered network with synchronization capability. In the proposed neural model, selection of the relevant information and rejection of irrelevant information for moving objects is achieved by modulating the temporal correlation of the neural activity (Kazanovich & Borisyuk, 2006; Mishra, Fellous, & Sejnowski, 2006) in two interconnected



**Fig. 12.** The tracking results of the algorithm on seven object tracks in PETS 2001 dataset. Coverage and  $F$  measure performance measures are shown for both surround and no surround cases. Surround processing improves coverage by 57% and  $F$  measure by 20%.

layers of retinotopically arranged neurons. In Layer 1, which represents the first interaction of low-level visual signals with a common (common to Layer 1 and Layer 2) oscillatory clock signal, inhibitory signals from Layer 2 modulate the neural activity produced by the distractor objects to become out-of-phase with respect to the clock. De-synchronization of the neuronal firing in Layer 1 prevents the formation of activity in the corresponding region of Layer 2. Thus, in Layer 2, which represents the outcome of attentional processing of moving objects, representations of unattended objects are “erased”. This mechanism of attentive tracking comes with a limitation, i.e. the inhibitory interactions in the surrounding regions of an attended object also cause interference among separate simultaneously attended regions. In other words, parameters of the inhibitory surround interactions from Layer 2 to Layer 1 define the spatial resolution of attentive tracking. The model nevertheless highlights the importance of oscillations and synchronization in neural information processing. In the sections below we seek to relate known physiology to various processing elements in the model.

### 4.1. Oscillations

The oscillations are suggested to originate from an interplay between intrinsic neural properties and network connectivity (Somers & Kopell, 1993; Steriade, 2001; Wang, 2010; Whittington & Traub, 2003). It has been shown that rhythmic firing of a neuron is phase locked with the membrane potential oscillations (Lampl et al., 1999; Lampl & Yarom, 1993; Volgushev, Chistiakova, & Singer, 1998). The relative time of the synaptic input with respect to the phase of the neuron’s oscillating membrane potential is critical for eliciting a spike. Oscillations in the membrane potential are suggested to be important for rhythmic and precise firing of neurons. Oscillatory depolarization in membrane potential normally produces repeated spikes with small inter-spike temporal jitters whereas sustained depolarization generates fewer spikes with greater inter-spike jitter. This is because cyclic hyperpolarization resets the inactivation  $\text{Na}^+$  currents, accelerating reactivation of  $\text{Na}^+$  channels (Volgushev et al., 1998). Thus, oscillatory depolarization creates a narrow time window of excitability for the neuron to accumulate synaptic input, and hence, facilitates precision in neural spiking (Lampl et al., 1999; Lampl & Yarom, 1993). However, it should be noted that, in general single neuron behavior is stochastic and irregular (Shadlen & Newsome, 1994; Softky & Koch, 1993). The oscillatory behavior in our model is exaggerated; future studies will address this and develop more accurate single neuron firing behaviors.

## 4.2. Synchronization

In a network of neurons in which neurons are not directly coupled, synchrony may emerge through common correlated input (Galán, Fourcaud-Trocm, Ermentrout, & Urban, 2006). The membrane potentials of visual cortical neurons with overlapping receptive fields synchronize even in the absence of firing and this synchronization is suggested to be due to extrinsic factors which are common to all synchronizing neurons (Lampl et al., 1999). Intracellular recordings have shown that membrane potential fluctuations are correlated with Local Field Potentials (LFPs) in visual cortex (Ferster & Carandini, 1996). Tukker, Fuentealba, Hartwich, Somogyi, and Klausberger (2007) suggest a causal relationship between firing pattern of GABAergic interneurons and LFP. Therefore, LFP can represent a clock signal for a population of neurons to synchronize their membrane potentials. Radman et al. (2007) showed that hyperpolarizing LFP delayed action potential while depolarizing LFP advanced it. The correlation between the LFPs of various distinct regions in cortex indicates that a common clock signal might be available to distinct neural populations (Fries et al., 1997; Siapas et al., 2005). However, it should be mentioned that LFPs are accumulated electrical activities from a population of neurons; hence LFP is created by coherent neural spiking. In order to solve this dilemma, we suggest that LFP and synchronization emerge together, but not one after another. However, there is evidence for prefrontal initiation of attention induced gamma oscillations, which travels down to sensory cortex (Gregoriou, Gotts, Zhou, & Desimone, 2009).

## 4.3. Oscillations and neural firing

How do the membrane potential oscillations affect neural firing? Lampl and Yarom (1993) injected current to elicit oscillations in the membrane potential of an olivary nucleus neuron, and induced synaptic activity at different phases with respect to the oscillation cycle of the membrane potential. They showed that the neuron does not linearly sum the oscillations and the synaptic activity, but there are supra and super linear operations depending on the phase of the synaptic input. The synaptic activity is amplified during the supra-linear cycle of the oscillations, evoking a spike (Gray & Singer, 1989). The cumulative effect of the membrane potential oscillations and synaptic input on neural firing is nonlinear, and we implemented this nonlinearity as multiplication in our model. Multiplication is suggested to be one of the most common non-linear operations realized in the nervous system (for a review, Koch & Segev, 2000). Radman et al. (2007) investigated the effect of Local Field Potentials (LFPs) on spike timing and found that hyperpolarizing LFP delayed action potential while depolarizing LFP advanced it. In the model, an oscillatory signal representing LFP is crucial in modulating the phase of the neural spikes.

The most crucial assumption of the model is using an external oscillatory modulation signal in the network instead of using interconnected inhibitory and excitatory neurons for generating oscillatory firing (e.g. Börgers et al., 2005). This simplification is validated by the findings in physiology literature mentioned above and it is necessary for studying the mechanisms in a complex network with feedforward and feedback connections. Specifically our model uses simple single neuron dynamics (LIF), however the model behavior is similar to models with more complicated neuron dynamics. Our model resembles the “soma-dendritic interference” model by Kamondi, Acsady, Wang, and Buzsaki (1998). In order to capture the neural dynamics in CA1 pyramidal neurons and phase coding in hippocampus, Kamondi et al. simulated a two compartment (soma and dendrite) neuron model. They applied dendritic and somatic sinusoidal input currents to the neuron

and observed that the phase of the neural firing with respect to local field oscillations depended on the magnitude of the dendritic synaptic input (Kamondi et al., 1998, Fig. 12). Similar dynamics are responsible in our model for the phase modulation of the neurons with respect to local field potential. In our model, the magnitude of the input to Layer 1 neuron is given by (see Appendix A):

$$A(t) * G(t) * \sin(2\pi ft),$$

where  $A(t)$  is sensory input and  $G(t)$  is the gain term due to feedback and horizontal connections. The  $G(t)$  term modulates the magnitude of the sinusoidal input and determines the phase of the neural firing with respect to the local oscillations.

Specifically, when a neuron in Layer 1 receives a large inhibitory feedback signal, it fires  $180^\circ$  out of phase with the local oscillation. This is because its net gain becomes negative, causing the neuron to integrate during the negative cycles of the clock. This behavior is the main reason why distractor neurons in close proximity to targets have enhanced representation in Layer 1, and this simulation result is the main drive for developing an object tracking algorithm with surround processing. Therefore it is important to point possible cortical mechanisms in support for this behavior although phase reversal ( $90^\circ$  shift) caused by “negative gain” is only a mathematical abstraction. Even though there is no direct evidence for the mentioned phase reversal, it is possible and actually probable. There are two arguments in support for this behavior:

1. The neuron is modeled with integrate and fire equations, however more complicated behavior emerges when multi-compartment Hodgkin–Huxley models are utilized, as explained above (Kamondi et al., 1998). In our model, there are two phases of neural firing (in-phase and  $180^\circ$  out-of-phase) which is determined by the magnitude of the sine input. Kamondi et al.’s model is able to shift the phase of neural firing in arbitrary amounts by modulating the magnitude of the sine input.
2. Inhibitory cells are affected more by attention than excitatory cells (Mitchell, Sundberg, & Reynolds, 2007) and it is shown in cat cortex that reduction in inhibition precedes neural spikes (Hasenstaub et al., 2005; Rudolph, Pospischil, Timofeev, & Destexhe, 2007). Therefore in a network of neurons biased by attention, “selected neuron” spikes are locked to the local inhibitory rhythm (see Tiesinga et al., 2008 for a review). In the cortex, inhibitory attentional feedback is expected to suppress the inhibitory interneurons and delay their firing. It is possible that this modulation entrains the neurons in the attentional surround for a half cycle shifted oscillatory rhythm. And integration of neural activity during the negative cycles is equivalent to half cycle shifted oscillatory drive signal in the model equations. A thorough literature review and communication with relevant papers’ authors concluded that there is no direct physiological study to confirm or reject the proposed effect of attention on distractor neuron activities.

## 4.4. Attention and synchrony

Attention and alertness are necessary conditions for coherence of LFPs (and neural spikes) in the gamma range across distant brain regions (Fries et al., 1997). Fries et al. argue that, attention might be the modulating the synchronicity of LFPs and spikes in the gamma range to increase the effectiveness of neural activity in the subsequent processing stages. EEG studies support the hypothesis of such long-range synchronization of neural activities due to top-down cognitive factors, by showing that top-down modulation enhances gamma-band EEG and coherence between electrodes (Keil, Gruber, & Muller, 2001).

De-synchronization of the neural activity of distractor stimulus has been shown in MEG studies and has been hypothesized as a mechanism to suppress irrelevant information (Gross et al., 2004). Reduced synchronization of neural activity for distractors is proposed to be due to inhibitory modulations of attention (Fries et al., 2001; Varela, Lachaux, Rodriguez, & Martinerie, 2001). In a more general scheme, the phase of the neural activity with respect to the common gamma oscillations was suggested to encode the saliency of the sensory inputs (for a review see, Fries et al., 2007). Even though temporal modulation of neural activity was suggested to be related to attentional selection in many studies, a biologically plausible feedforward/feedback model that is based on hierarchical processing of cortex was not proposed previously to our knowledge.

#### 4.5. Cortical delays

Delays between excitatory and inhibitory connections have been shown in cortex (Carandini, Heeger, & Senn, 2002; Monier, Chavane, Baudot, Graham, & Frégnac, 2003). Recently, Sundberg et al. (2009) showed that attentional surround suppression is more delayed compared to enhancement. In our model, differential delay of excitatory and inhibitory attentional top-down modulations is crucial in producing out-of-phase/in-phase firings observed in distractor/target neurons. Also it should be noted that, the period of oscillations in the model is twice the feedback delay. A similar dynamic was observed in another modeling study (Brunel & Hakim, 1999), which analytically analyzed sparsely connected networks. The oscillations emerged in the network when the feedback is strong, which is also observed in the model presented here (not shown).

#### 4.6. Inhibitory surround of attention

The attentional modulation of the activity of neurons responding to a distractor depends on the distance between the target and the distractor in single-unit recordings (Gawne & Martin, 2002; Luck et al., 1997). Recent physiological studies on the spatio-temporal characteristics of attentional window suggest an inhibitory surround of attentional modulation (Hopf, Boehler, Luck, Heinze, & Schoenfeld, 2006). The inhibitory surround of attention is suggested to suppress distractor objects in the attentional network (Moore & Armstrong, 2003; Schall & Hanes, 1993). The suppression filters out the irrelevant inputs, such that the attention-related visual areas are activated only by the relevant target objects (Everling et al., 2002; Kastner & Pinsky, 2004; Schall & Thompson, 1999). Cortical recordings have shown that distractors do not interfere with targets in extrastriate visual cortex (Chelazzi et al., 1998; Everling et al., 2002; Recanzone & Wurtz, 2000; Reynolds, Chelazzi, & Desimone, 1999) and the mechanism for the reduction of distractor interference is suggested to be attention-related feedback from the attentional network onto the extrastriate cortex (Everling et al., 2002; Kastner & Pinsky, 2004; Schall & Thompson, 1999). Therefore the feedforward–feedback loop between extrastriate areas and the attention-related network realizes the selection process of information of interest, while ignoring the rest. Although the surround of attention suppresses distractors, it also causes interference among the targets when multiple objects are attended (Bahcall & Kowler, 1999; Cutzu & Tsotsos, 2003; Skelton & Eriksen, 1976). When a target is inside the inhibitory attentional zone of another target, the net processing strength should be reduced. Therefore the attentional surround poses a limit on the total number of attended regions, although it serves to remove interference from distractors and sharpens the contrast between relevant and irrelevant information. The model proposed here implements the

attention-related temporal modulation of neural activity with a local modulatory surround of attentional influence on low-level visual processing.

How does attentional surround affect low level visual areas? Even though there is no direct study, Maier et al. (2008) showed that V1 activity may survive even for perceptually suppressed stimuli. Therefore, low level visual features are computed even though the stimulus is not consciously registered. In our model, attentional surround triggers activity in the first layer of the network hence initiates low level feature processing however this activity does not propagate to second layer (and distractor stimuli are not attentively tracked). We are suggesting that enhancing activity of background stimuli in low level feature processing neural substrate is essential for correct correspondence assignments. By using surround processing in a feature based object tracking algorithm, we show that target features can be more effectively extracted: good features that are dissimilar to background or distractor can be computed.

#### 4.7. Attentional network

Brain imaging studies of posterior parietal cortex suggest that it plays a role in visual attention (Kanwisher & Wojciulik, 2000; Wojciulik & Kanwisher, 1999) and the attentive tracking of objects (Culham et al., 1998; Jovicich et al., 2001). Activity in parietal areas is suggested to reflect attentional processing rather than sensory visual information coming from low level brain areas. In addition, studies on MOT suggest that attentional windows (spotlights) are able to follow objects that are in motion (Cavanagh & Alvarez, 2005). Physiological studies suggest feedforward–feedback loop between the early visual areas and the higher attention-related areas (Vidyasagar, 1998). Correlation between the LFP in early processing and attention-related areas during an attention demanding task (Fries et al., 1997) suggest the existence of a common clock that synchronizes the activities from these two distinct regions in the brain (Engel et al., 2001). Based on the evidence, the model adopts a two-layer structure with inter-layer and intra-layer connections for attentional processing. In the model, the tagging process activates Layer 2 neurons and initiates the attentional window that follows the target as it moves. The synchronous activity of target neurons with the common clock signal provides attentional selection and facilitates further visual processing (not modeled). Whereas, out-of-phase firing of distractor neurons provides a mechanism to suppress irrelevant information from visual awareness, as binding will not be achieved due to an incoherent temporal pattern with respect to the common clock (Fries et al., 2007).

#### 4.8. Other modeling studies

Selection of relevant visual input via temporal modulation of the neural representation is a powerful concept that is being used in recent models of attention (Börger et al., 2005; Buia & Tiesinga, 2006; Kazanovich & Borisyuk, 2006; Mishra et al., 2006; Olufsen et al., 2003; Tiesinga, Fellous, Salinas, Jose, & Sejnowski, 2004; Tiesinga & Sejnowski, 2004). Kazanovich and Borisyuk (2006) used phase oscillators as neural units and built a multilayer network that is based on phase locking, resonance and adaptation to implement interaction between oscillators. Visual objects are represented by synchronized assemblies of oscillators, and the attended objects showed synchronous activity with a central oscillator. A separate layer of neural oscillators and a separate central oscillator were assigned for each attended object, and desynchronizing connections was used to segregate the visual objects (that are in different layers) in frequency space. Kazanovich and Borisyuk showed that, the proposed neural



network can successfully track 5 visual objects, consistent with the psychophysical data. The main difference between their model and the model we have proposed in this paper is that a separate layer of network and a central oscillator was implemented for each target object in Kazanovich and Borisjuk's model whereas our model has a single layer for all the attended and non-attended objects. Also, in their model desynchronization was achieved via connections between the central oscillators of each layer, but in our model it is a product of the inhibitory surround of attentional feedback. More physiological, psychophysical and modeling studies are needed to elucidate the details of attentional mechanisms and their influence on the timing of neural activity.

Börgers and Kopell (2003) investigated the interplay between excitatory and inhibitory cells to produce synchronization. Their model suggest that a population of neurons can show synchronization when they receive “a common excitatory synaptic input pulse”. The interplay between excitatory and inhibitory interneurons can be responsible for the emergent clock signal which causes synchrony in Layer 2 of our model when a stimulus is attended. A series of studies investigated the characteristics of inhibitory networks under attentional bias (Buia & Tiesinga, 2006; Tiesinga et al., 2004; Tiesinga & Sejnowski, 2004). The main hypothesis of these studies is that, selective attention increases the firing rate of a subset of interneurons and synchronizes the neurons representing the selected stimulus. The simulations have shown the importance of phase with respect to the inhibitory rhythm in transmitting the activity to other connected networks in a feedforward manner. However, the behavior of an interconnected network with feedforward and feedback projections is not investigated. Mishra et al. have also proposed a feedforward model to account for response properties of V4 neurons when they are presented with two differentially preferred stimuli. They concluded that an inhibitory phase shift mechanism is used by top-down attention to bias neural activity.

The level of abstraction and mechanisms of the models mentioned above and of our model are very different, as we want to specifically investigate the attentional surround mechanisms:

1. Our model adopts feedforward–feedback connections for phase modulation.
2. Oscillatory signal multiplicatively modulates the input gain.
3. Feedback inhibition is more delayed than excitation that causes out-of-phase firing of distractor neurons in the surround.
4. Distractor activity is enhanced by attentional surround that has clear advantages for energy efficiency and correspondence problem, as shown in Section 3.4.

The phase modulation of neural activity is investigated in a number of studies (see Wang, 2010 for a review) by injecting synaptic input with 3 components: a baseline, a small sinusoidal wave and noise. “soma-dendritic interference” model by Kamondi et al. (1998) discussed above and the model proposed in this paper are in this category of studies where synchrony is a product of common input. Future studies will explore the behavior of the proposed model with more realistic single neuron dynamics.

#### 4.9. The object tracking algorithm

The Object tracking algorithm showed the improvement in tracking with surround processing. Surround processing enables feature extraction in the surround and these features are used to eliminate target features (inside the tracking window) that are likely to be miscomputed due to appearance similarities. The intention in developing an object tracking algorithm with surround processing is not to create a state-of-the-art object tracker but to explore and contribute to a class of algorithms that are surround conscious. Also, although there are analogies between

the neural network results and the surround processing approach implemented in an object tracking algorithm, they are not strong.

Chen and Yang (2007) and Han, Ye, and Jiao (2011) also exploited surround information but their approach and mine have important differences:

1. Chen and Yang's is a region based tracking algorithm and Han et al.'s is a dense patch based tracking algorithm, but mine is feature-based.
2. Surround information is used to weigh target regions or patches in two approaches, but it is used to reject target features in my algorithm. So my approach resembles winner-take-all through lateral inhibition.
3. In their algorithm both studies only calculate the similarity metric between target regions and surround regions. But in my algorithm for evaluating the probability of correspondence mistakes of each feature, the similarity metric is calculated between the target feature and all surround features, as well as between the target feature and other target features. This way, incorrect correspondences between target features are avoided. Therefore my algorithm resembles lateral inhibition between features, and inhibition does not discriminate between target and surround features.
4. A segmentation step is necessary for Chen and Yang's algorithm but feature extraction is performed for my algorithm. Feature extraction is known to focus on locations that have high information content, i.e. corners. However a segmented region might have uniform intensity which is problematic for matching. Han et al.'s algorithm computes a histogram based feature for every pixel which is more prone to uniform intensity region problems than Chen and Yang's approach. Hence, the algorithm presented in this paper exploits the salient location detection concept while previous work do not.
5. Similarity metric is color histogram in Chen and Yang's algorithm but it is patch correlation in mine which is expected to be more discriminative. Han et al. uses a very high dimensional feature for every pixel in the target region.
6. Han et al. uses the similarity metric of pixels in a particle filter framework which shows similarities to a swarm intelligence based tracking, however particle filters are computationally much more demanding.
7. In general the two algorithms are computationally more demanding than mine since the algorithms search segmented regions or pixel features in the following frame, but my algorithm has to find correspondences for sparse features.

There is room for improvement in my algorithm by extracting more information, (such as color, gradient, motion energy etc.) from the surround and refining feature extraction in the tracking window. However it should be noted that the algorithm presented here has closer resemblance to cortical processing in human visual system than previous studies.

Compared to other swarm intelligence approaches in object tracking, our swarm tracker does not require offline training (Kölsch & Turk, 2005) and is based on simple statistics of swarm members as opposed to complex interactions between swarm members (Canalis, Sanchez-Nielsen, & Hernandez-Tejera, 2006). The interaction between the swarm members occurs during swarm spread phase for preventing accumulation around a strong corner feature. Also the killing of a swarm member is decided by the motion of other members, which is an indirect interaction. The gradual elimination of unsuccessful members and aperiodic respread of swarm not only stabilizes tracking but also immunizes the algorithm to size and appearance changes, and this behavior is novel in swarm intelligence tracking approaches.



## 5. Conclusion

Our modeling studies suggest that the neural activity representing unattended stimuli can be modulated in order to temporally tag them as irrelevant. This modulation is expected to prevent these stimuli to be processed further by higher order visual areas. Our model emphasizes the importance of feedback surround and its phase with respect to the local neural activity in the network. Surround signals can cause interference between targets in close proximity, but also enhance the feature processing for distractors that are in close range to the targets. This type of surround processing (i.e. enhancement of distractor feature processing) is implemented in an object tracking algorithm and it improved tracking performance by more than 50%. Compared to other object tracking algorithms that are surround conscious, the algorithm presented here has closer resemblance to cortical processing in human visual system due to its roots in lateral inhibition, winner-take-all behavior and saliency detection. In the future the surround processing based object tracking algorithm will be implemented using neural networks.

## Acknowledgments

I want to thank Haluk Ogmen, and Saumil S. Patel for their invaluable help in this study. I also want to thank anonymous reviewers for their very important input.

## Appendix A. Model equations

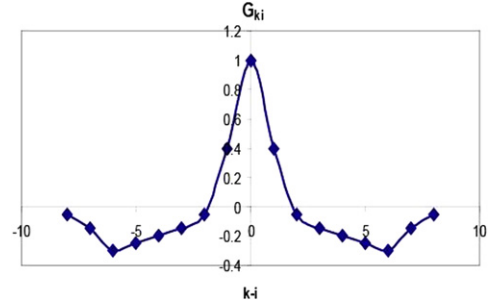
The difference equation of a layer 1 neuron residing at location  $i$  is

$$\begin{aligned} u_i(t) - u_i(t-1) &= 10^{-5} * \text{Noise}(t) + 10^{-4} * \left[ (-u_i(t-1) - V_{\text{rest}})/\eta \right. \\ &+ \left\{ 1 + \text{Gain}_{fb} \sum_{\tau=0}^{4\lambda} e^{-\tau/\lambda} \sum_{k=i-8}^{k=i+8} G_{ki} Y_{i+k}(t-\tau-\delta_{fb}) \right. \\ &+ \left. \left. \text{Gain}_h \sum_{\tau=0}^{4\lambda} e^{-\tau/\lambda} (U_{i-1}(t-\tau-\delta_h) + U_{i+1}(t-\tau-\delta_h)) \right\} \right. \\ &\left. * \text{drive}(t) * \text{Input}(t) \right] \end{aligned}$$

and the sinusoidal signal is given by  $\text{drive}(t) = 1.25 * \sin(2 * \pi * 35 * 10^{-4} * t)$ .

The small letters represent the membrane potential of the neurons while the capital letters represent the action potential patterns. The neuron gives a spike of amplitude 1 when the activity ( $u_i$  or  $y_i$ ) exceeds  $\text{threshold} = 0.02$ . After the spike there is a *refractory period* = 20 simulation units during which the neuron cannot integrate. After the refractory period, the neuron starts to integrate activity again.

The  $\text{Noise}(t)$  term is a random number uniformly distributed between 0 and 1. The neuron leaks some of its activity via the third additive term in the first line. The term in the curly parentheses is the overall input gain of the neuron.  $\text{Input}(t)$  is the sensory input to the neuron. There are two factors that modulate the input gain: feedback from layer 2 and horizontal connections.  $\text{Gain}_{fb}$  determines the strength of the feedback connections and  $\text{Gain}_h$  the strength of the horizontal connections. A sinusoidal drive input multiplies the input causing oscillatory bursts of the same frequency as the drive signal. The summation over time represents the synaptic integration after an action potential, where  $\lambda$  is the



**Fig. 13.** The receptive field profile of  $G_{ki}$  that determines the spatial distribution of the feedback from the second layer. The delay from layer 2 to layer 1 for positive values of  $G_{ki}$  is  $\delta_{fb}^+$  and for negative values of  $G_{ki}$  is  $\delta_{fb}^-$ .

synaptic halftime. The synaptic integration starts from  $\tau = 0$  and runs up to  $\tau = 4\lambda$ .  $Y_{i+k}$  is the spiking activity of a neuron in layer 2 that modulates the input gain of neuron  $u_i$ . The feedback is summed over space and time. The spatial distribution of the feedback from the second layer is determined by a receptive field profile of  $G_{ki}$  (Fig. 13). There is a feedback delay from layer 2 to layer 1 which is represented by  $\delta_{fb}$ . The value of  $\delta_{fb}$  is different for positive and negative values of  $G_{ki}$  because it is assumed that the inhibitory feedback has longer delay than excitatory feedback. This dichotomy will be depicted by  $\delta_{fb}^+$  for positive feedback and  $\delta_{fb}^-$  for negative feedback.  $U_{i-1}$  and  $U_{i+1}$  are neighboring neurons that send facilitatory feedback to  $u_i$  and modulate its input gain. There is also a delay in these horizontal connections denoted by  $\delta_h$ .

The difference equation of a layer 2 neuron residing at location  $i$  is

$$\begin{aligned} y_i(t) - y_i(t-1) &= 10^{-5} * \text{Noise}(t) + 10^{-4} * \left[ (-y_i(t-1) - V_{\text{rest}})/\eta \right. \\ &+ \left\{ 1 + \text{Gain}_{h2} \sum_{\tau=0}^{4\lambda} e^{-\tau/\lambda} (Y_{i-1}(t-\tau-\delta_{h2}) \right. \\ &+ \left. \left. Y_{i+1}(t-\tau-\delta_{h2}) \right) \right\} * \text{drive}(t) * \text{Input}_{ff}(t) \right]. \end{aligned}$$

The input is given by

$$\text{Input}_{ff}(t) = \text{Gain}_{ff} \sum_{\tau=0}^{4\lambda} e^{-\tau/\lambda} \sum_{k=i-1}^{k=i+1} H_{ki} U_{i+k}(t-\tau-\delta_{ff}).$$

The equation is similar to the one above, but for the layer 2 neuron there is a feedforward input from layer 1 neuron  $U_{i-k}$ , instead of a sensory input.  $\text{Gain}_{ff}$  for the strength of the feedforward input and there is a receptive field profile  $H_{ki}$  that determines the spatial distribution. The feedforward input is delayed by the amount  $\delta_{ff}$ . The horizontal connections from neighbor neurons  $Y_{i-1}$  and  $Y_{i+1}$  modulate the feedforward input gain of the neuron. The strength of the horizontal connections is determined by  $\text{Gain}_{h2}$  and they are delayed by  $\delta_{h2}$ .

The parameters of the equations are given in [Box I](#).

## Appendix B. The object tracking algorithm

The Object tracking algorithm detects several corners in the target region and tracks the corners using template matching (normalized cross correlation). The individual corner tracks are brought together to infer the target motion direction, speed and size. The corners that are tracked are refreshed from time to time

```

Vrest = 0.01
η = 0.02
Gainfb = 43
Gainh = 4
Gainh2 = 14
Gainff = 120
Gki = [-0.05 -0.15 -0.3 -0.25 -0.2 -0.15 -0.05 0.4 1 0.4 -0.05 -0.15 -0.2 -0.25 -0.3 -0.15 -0.05]
Hki = [0.02 1 0.02]
δfb+ = 140
δfb- = 280
δh = 280
δh2 = 280
δff = 140
λ = 20

```

#### Box 1.

in order to cope with size changes. There are three phases of the algorithm (Section 2.2.1. and Fig. 2). The flow chart of each of these phases is given in Fig. 14. The details of phases are explained below. Surround processing is a part of initialization and swarm update phases and it can be turned on or off. It is explained in a separate section below. There are a few parameters in the algorithm, for that reason they are not named but embedded in the formulas.

#### B.1. Initialization

When the target location and window size are specified (Fig. 14(A)), the corner magnitudes in the target region are calculated in the image from phase congruency (Kovesi, 1999, 2003). The number of tracking windows initiated on the target depends on the target size. In the image set we tested, number of tracking windows was always 24. When the number of tracking windows needed is decided,  $16 \times 16$  size correlation tracking windows are initiated on the strongest corners after non-maxima suppression as shown in Fig. 2. Non-maxima suppression is assigning zero corneriness to the pixels that are in the neighborhood of a “selected” strong corner. This prevents accumulation of tracking windows on one strong cluster of corners, and provides a balanced spread of the swarm over the target. Non-maxima suppression eliminates the corners in the immediate neighborhood of a selected corner and the extent of the neighborhood is adaptively changed with size of the target window:

$$\text{WidthNeighborhood} = \text{Width}/19.6;$$

$$\text{HeightNeighborhood} = \text{Height}/19.6.$$

#### B.2. Tracking

The reference image patches from 24 tracking windows are tracked separately in each frame using normalized cross correlation (NCC) (Fig. 14(B)) and it is given by:

$$\text{NCC} = \frac{1}{n-1} \sum_{x,y} \frac{(p_r(x,y) - \bar{p}_r)(p(x,y) - \bar{p})}{\sigma_r \sigma}$$

in which the summation is over the size of the patches.  $p_r$  is the reference image patch from first frame,  $p$  is the image patch from a subsequent frame. Means are subtracted from the patches and the result is divided by the standard deviation in order to be immune

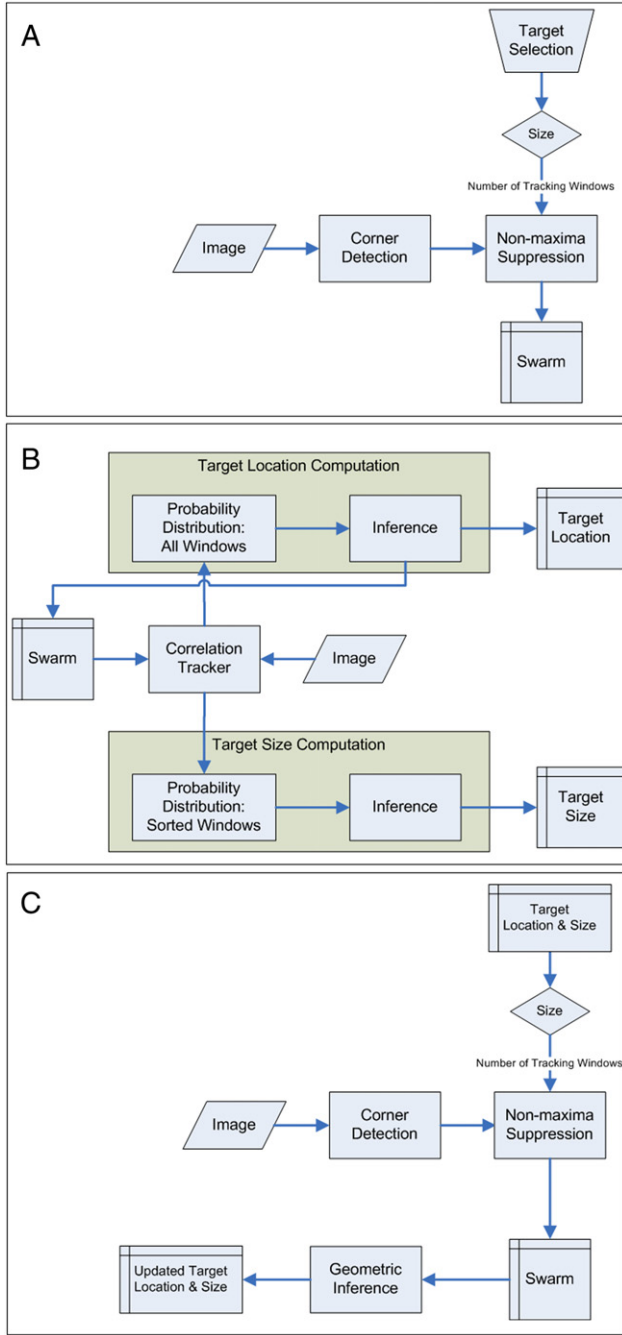
to intensity changes. NCC is calculated in  $32 \times 32$  pixels search window by shifting the  $p$  image patch and the pixel coordinates with highest NCC is assigned as the new target location in that frame.

Tracking can be decomposed into the estimation of target location and size. For target location estimation, the distribution of the motion of each tracking window from frame to frame is calculated. This statistics generates an estimate of collective motion, which is due to the actual target motion. Using this statistics, algorithm is able to reject outlier tracking window motion and arrives at a robust estimate of target motion. The tracking windows which have motion inconsistent with the swarm motion are considered as outliers and dropped, hence swarm is refined. In order to reject outliers, standard deviation of the motion for  $x$  and  $y$  axes are computed. The tracking windows that are  $2 * \text{std} + 0.5$  pixels away from the mean motion are considered as outliers and tracking is killed in the next frame. The target motion is calculated from the mean motion of inlier tracking windows.

The swarm should be able to adjust to changes in target size which is critical for swarm update phase. In order to satisfy this requirement, the statistics of tracking window motion is examined to extract information about size changes. The tracking windows are sorted according to their coordinates (windows are grouped as being on the left and right sides for  $x$  axis and bottom and top sides for  $y$  axis) and the associated motions of these tracking windows are listed with the same resulting order. Then, size change for each axis is computed from the motion of these two tracking window distributions (left vs. right and bottom vs. top). For example, in case of a size change in  $x$  axis, mean value of the left tracking window motion is expected to be different from the mean value of the right tracking window motion. For an increase of size in  $x$  axis, tracking windows on the left side of the target will have negative sign of motion whereas the right ones will have positive sign of motion. The difference between the two distributions is an estimate of the change in size, which is purely obtained from the analysis of swarm motion profile.

#### B.3. Swarm update

As outlier tracking windows are dropped, the number of members in the swarm decreases in time. If it decreases below a threshold (19), swarm is re-spread on the target area defined by the center of the swarm and the size of the target (Fig. 14(C)). In this update phase, the corners are calculated and tracking window locations are computed after non-maxima suppression as it is done in the initialization phase. Swarm update phase is very similar



**Fig. 14.** The flowcharts of initialization, tracking and swarm update phases of the object tracking algorithm shown in (A), (B) and (C) respectively.

to initialization phase. However in addition to the processes in the initialization phase, center location and the size of the swarm is updated according to the spread of the swarm. Thus, swarm update phase utilizes the distribution of selected corners in order to contribute to the center and size estimation. This stage is called geometric inference since it processes geometric information of the swarm to arrive at size and location estimates of the target. In geometric inference, mean and standard deviation of the corner locations are used to update the center of the target:

$$\text{CenterNew} = \text{mean}(\text{WindowLocations});$$

$$\text{StdSwarm} = \text{std}(\text{WindowLocations});$$

$$\text{CenterChange}$$

$$= 1.5 \times (\text{CenterOld} - \text{CenterNew}) \times (\text{StdSwarm}/\text{SizeTarget});$$

whereas min–max of the corner locations is used to update size of the target:

$$\text{SizeTargetNew}$$

$$= \max(\text{WindowLocations}) - \min(\text{WindowLocations}) + 8;$$

$$\text{SizeChange} = 0.2 \times (\text{SizeTargetOld} - \text{SizeTargetNew}).$$

As seen above, alpha filters are used to apply center and size updates, hence center and size computed from appearance (corner locations) gradually influences the center and size of the target.

#### B.4. Surround processing

Surround processing feature is a part of initialization and swarm update phases, it can be turned on or off. It allows the selection of good corners for tracking. In order to achieve this, 24 corners are detected in the surround of the target region, which is double the size of the target region ( $2 \times \text{Width}$  by  $2 \times \text{Height}$ ).  $16 \times 16$  pixel window features are extracted from these surround corners. And 48 corners are detected in the target region which is twice the number when surround processing is turned off, but only half of them are selected after quality computation based on distance in feature space. Similarity distance between the target features and surround features are computed as follows:

$$\text{DistCorrelation}_{i,j} = \text{NCC}(\text{TargetFeature}_i, \text{SurroundFeature}_j);$$

$$\text{Dist}_{i,j} = \text{DistCorrelation}_{i,j} \times \text{DistEuclidean}(\text{TargetFeature}_i, \text{SurroundFeature}_j).$$

The overall distance between the features is the normalized cross correlation (NCC) of the features weighted by the Euclidean pixel distance. The distances are sorted and the best 24 target features are selected for tracking.

#### Appendix C. Supplementary data

Supplementary material related to this article can be found online at [doi:10.1016/j.neunet.2012.01.005](https://doi.org/10.1016/j.neunet.2012.01.005).

#### References

- Abeles, M. (1982). Role of the cortical neuron: integrator or coincidence detector? *Israel Journal of Medical Sciences*, 18(1), 83–92.
- Ardid, S., Wang, X. J., Gomez-Cabrero, D., & Compte, A. (2010). Reconciling coherent oscillation with modulation of irregular spiking activity in selective attention: gamma-range synchronization between sensory and executive cortical areas. *The Journal of Neuroscience*, 30, 2856–2870.
- Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research*, 39(1), 71–86.
- Bedell, H. E., Chung, S. T. L., Ogmen, H., & Patel, S. S. (2003). Color and motion: which is the tortoise and which is the hare? *Vision Research*, 43, 2403–2412.
- Bonabeau, E., Dorigo, M., & Thérault, G. (2000). Inspiration for optimization from social insect behavior. *Nature*, 406, 39–42.
- Börgers, C., Epstein, S., & Kopell, N. J. (2005). Background gamma rhythmicity and attention in cortical local circuits: a computational study. *Proceedings of the National Academy of Sciences USA*, 102, 7002–7007.
- Börgers, C., & Kopell, N. (2003). *Neural Computation*, 15, 509–539.
- Bringuier, V., Chavane, F., Glaeser, L., & Fregnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science*, 283(5402), 695–699.
- Brody, C. D., & Hopfield, J. J. (2003). Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron*, 37, 843–852.
- Brunel, N., & Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, 11, 1621–1671.
- Buia, C., & Tiesinga, P. (2006). Attentional modulation of firing rate and synchrony in a model cortical network. *Journal of Computational Neuroscience*, 20, 247–264.
- Buzsaki, G., & Chrobak, J. J. (1995). Temporal structure in spatially organized neuronal ensembles: a role for interneuronal Networks. *Current Opinion in Neurobiology*, 5(4), 504–510.
- Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304, 1926–1929.
- Canalis, L. A., Sanchez-Nielsen, E., & Hernandez-Tejera, M. (2006). Swarmtrack: a particle swarm approach to visual tracking. In *VISAPP 06, Vol. 2* (pp. 197–200).
- Carandini, M., Heeger, D. J., & Senn, W. (2002). A synaptic explanation of suppression in visual cortex. *The Journal of Neuroscience*, 22(22), 10053–10065.

- Castelo-Branco, M., Neuenschwander, S., & Singer, W. (1998). Synchronization of visual responses between the cortex, lateral geniculate nucleus, and retina in the anesthetized cat. *The Journal of Neuroscience*, 18(16), 6395–6410.
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9(7), 349–354.
- Chelazzi, L., Duncan, J., Miller, E., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, 80, 2918–2940.
- Chen, D., & Yang, J. (2007). Robust object tracking via online dynamic spatial bias approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319–349.
- Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, 80, 2657–2670.
- Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vision Research*, 43, 205–219.
- Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2, 704–716.
- Everling, S., Tinsley, S. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience*, 5(7), 671–676.
- Ferster, D., & Carandini, M. (1996). Spontaneous fluctuations in membrane potential of complex cells in visual cortex. *Society for Neuroscience Abstracts*, 22, 490.
- Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, 32, 209–224.
- Fries, P., Nikolic, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, 30(7), 309–316.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291, 1560–1563.
- Fries, P., Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proceedings of the National Academy of Sciences USA*, 94, 12699–12704.
- Galán, R., Fourcaud-Trocm, N., Ermentrout, G., & Urban, N. (2006). Correlation-induced synchronization of oscillations in olfactory bulb neurons. *The Journal of Neuroscience*, 26, 3646–3655.
- Gawne, T. J., & Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *Journal of Neurophysiology*, 88, 1128–1135.
- Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences USA*, 86(5), 1698–1702.
- Gregoriou, G. G., Gotts, S. J., Zhou, H., & Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*, 324, 1207–1210.
- Grinvald, A., Lieke, E. E., Frostig, R. D., & Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, 14, 2545–2568.
- Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., et al. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the National Academy of Sciences USA*, 101, 13050–13055.
- Han, Z., Ye, Q., & Jiao, J. (2011). Combined feature evaluation for adaptive visual object tracking. *Computer Vision and Image Understanding*, 115, 69–80.
- Hasenstaub, A., et al. (2005). Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron*, 47, 423–435.
- Hopf, J. M., Boehler, C. N., Luck, S. J., Heinze, H. J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences USA*, 103(4), 1053–1058.
- Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*, 376, 33–36.
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2, and V3 neurons. *Nature*, 394, 784–787.
- Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *Journal of Cognitive Neuroscience*, 13, 1048–1058.
- Kamondi, A., Acsády, L., Wang, X. J., & Buzsáki, G. (1998). Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: activity-dependent phaseprecession of action potentials. *Hippocampus*, 8, 244–261.
- Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Review Neuroscience*, 1, 91–100.
- Kastner, S., & Pinsk, M. A. (2004). Visual attention as a multilevel selection process. *Cognitive, Affective, & Behavioral Neuroscience*, 4(4), 483–500.
- Kazanovich, Y., & Borisyuk, R. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation*, 18(6), 1413–1440.
- Keil, A., Gruber, T., & Müller, M. M. (2001). Functional correlates of macroscopic high-frequency brain activity in the human visual system. *Neuroscience & Biobehavioral Reviews*, 25(6), 527–534.
- Koch, C., & Segev, I. (2000). The role of single neurons in information processing. *Nature Neuroscience Supplement*, 3, 1171–1177.
- Koepsell, K., & Sommer, F. (2008). Information transmission in oscillatory neural activity. *Biological Cybernetics*, 99, 403–416.
- Kölsch, M., & Turk, M. (2005). Hand tracking with flocks of features. In *IEEE computer society conference on computer vision and pattern recognition*. Vol. 2 (p. 1187).
- König, P., Engel, A. K., Roelfsema, P. R., & Singer, W. (1995). How precise is neuronal synchronization? *Neural Computation*, 7(3), 469–485.
- König, P., Engel, A. K., & Singer, W. (1996). Integrator or coincidence detector? the role of the cortical neuron revisited. *Trends in Neurosciences*, 19(4), 130–137.
- Kovesi, P. (1999). Image features from phase congruency. *A Journal of Computer Vision Research*, 1(3).
- Kovesi, P. (2003). Phase congruency detects corners and edges. In *The Australian pattern recognition society conference: DICTA 2003* (pp. 309–318).
- Lampl, I., Reichova, I., & Ferster, D. (1999). Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron*, 22, 361–374.
- Lampl, I., & Yarom, Y. (1993). Subthreshold oscillations of the membrane potential: a functional synchronizing and timing device. *Journal of Neurophysiology*, 70(5), 2181–2186.
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention to areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, 77, 24–42.
- Maier, J., et al. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature Neuroscience*, 11, 193–200.
- Mirollo, R., & Strogatz, S. H. (1990). Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics*, 50(6), 1645–1662.
- Mishra, J., Fellous, J. M., & Sejnowski, T. J. (2006). Selective attention through phase relationship of excitatory and inhibitory input synchrony in a model cortical neuron. *Neural Networks*, 19(9), 1329–1346.
- Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron*, 55, 131–141.
- Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron*, 63, 879–888.
- Monier, C., Chavane, F., Baudot, P., Graham, L., & Frégnac, F. (2003). Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. *Neuron*, 37(4), 663–680.
- Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421(6921), 370–373.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229, 782–784.
- Ogmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research*, 43, 1337–1350.
- Olufsen, M. S., Whittington, M. A., Camperi, M., & Kopell, N. (2003). New roles for the gamma rhythm: population tuning and preprocessing for the beta rhythm. *Journal of Computational Neuroscience*, 14, 33–54.
- Polyshyn, Z. W. (2006). Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Visual Cognition*, 14(2), 175–198.
- Radman, T., Su, Y., An, J. H., Parra, L. C., & Bikson, M. (2007). Spike timing amplifies the effect of electric fields on neurons: implications for endogenous field effects. *The Journal of Neuroscience*, 27(11), 3030–3036.
- Recanzone, G. H., & Wurtz, R. H. (2000). Effects of attention on MT and MST neuronal activity during pursuit initiation. *Journal of Neurophysiology*, 83, 777–790.
- Reynolds, J., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of Neuroscience*, 19, 1736–1753.
- Rudolph, M., Pospischil, M., Timofeev, I., & Destexhe, A. (2007). Inhibition determines membrane potential dynamics and controls action potential generation in awake and sleeping cat cortex. *The Journal of Neuroscience*, 27, 5280–5290.
- Schall, J. D., & Hanes, D. P. (1993). Neural basis of saccade target selection in frontal eye field during visual search. *Nature*, 366, 467–469.
- Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience*, 22, 241–259.
- Schlotterer, G. (1977). Responses of the locust descending movement detector neuron to rapidly approaching and withdrawing visual stimuli. *Canadian Journal of Zoology*, 55, 1372–1376.
- Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4, 569–579.
- Shi, J., & Tomasi, C. (1994). Good features to track. In *Proceedings of the conference on computer vision and pattern recognition*. (pp. 593–600).
- Siapas, A. G., Lubenov, E. V., & Wilson, M. A. (2005). Prefrontal phase locking to hippocampal theta oscillations. *Neuron*, 46, 141–151.
- Silver, M. A., Ress, D., & Heeger, D. J. (2005). Topographic maps of visual spatial attention in human parietal cortex. *Journal of Neurophysiology*, 94, 1358–1371.
- Skelton, J. M., & Eriksen, C. W. (1976). Spatial characteristics of selective attention in letter matching. *Bulletin of the Psychonomic Society*, 7(2), 136–138.
- Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of Neuroscience*, 13, 334–350.
- Somers, D., & Kopell, N. (1993). Rapid synchronization through fast threshold modulation. *Biological Cybernetics*, 68(5), 393–407.
- Steriade, M. (2001). Impact of network activities on neuronal properties in corticothalamic systems. *Journal of Neurophysiology*, 86, 1–39.
- Sundberg, K. A., Mitchell, J. F., & Reynolds, J. H. (2009). Spatial attention modulates center-surround interactions in macaque visual area v4. *Neuron*, 61, 952–963.
- Thompson, K. G., & Bichot, N. P. (2004). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, 147, 249–262.
- Tiesinga, P. H., Fellous, J. M., Salinas, E., Jose, J. V., & Sejnowski, T. J. (2004). Inhibitory synchrony as a mechanism for attentional gain modulation. *Journal of Physiology (Paris)*, 98, 296–314.



- Tiesinga, P., Fellous, J. M., & Sejnowski, T. J. (2008). Regulation of spike timing in visual cortical circuits. *Nature Reviews Neuroscience*, 9, 97–107.
- Tiesinga, P. H. E., & Sejnowski, T. J. (2004). Rapid temporal modulation of synchrony by competition in cortical interneuron networks. *Neural Computation*, 16, 251–275.
- Tukker, J., Fuentealba, P., Hartwich, K., Somogyi, P., & Klausberger, T. (2007). Cell type-specific tuning of hippocampal interneuron firing during gamma oscillations in vivo. *The Journal of Neuroscience*, 27, 8184–8189.
- Varela, F. J., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brain web: phase-synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2, 229–239.
- Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. *Neuroreport*, 9, 1947–1952.
- Volgushev, M., Chistiakova, M., & Singer, W. (1998). Modification of discharge patterns of neocortical neurons by induced oscillations of the membrane potential. *Neuroscience*, 83(1), 15–25.
- Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, 90, 1195–1268.
- White, T., & Pagurek, B. (1998). Towards multi-swarm problem solving in networks. In *Proc of the 3rd int. conf. on multi-agent systems*. (pp. 333–340).
- Whittington, M., & Traub, R. D. (2003). Interneuron diversity series: inhibitory interneurons and network oscillations in vitro. *Trends in Neurosciences*, 26(12), 676–682.
- Winfree, A. T. (1980). *The geometry of biological time*. New York: Springer.
- Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron*, 23, 747–764.
- Womelsdorf, T., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, 316, 1609–1612.
- Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: a survey. *ACM Computing Surveys*, 38(4), 1–45.
- Yilmaz, O., Tripathy, S. P., Patel, S. S., & Ogmen, H. (2007). Attraction of flashes to moving dots. *Vision Research*, 47, 2603–2615.